

Clasificación en tiempo real de materias primas mediante un sensor espectroscópico de imagen y técnicas lineales de procesamiento de datos

Real time raw material characterization by means of an imaging spectroscopic sensor and linear data processing techniques

P. Beatriz GARCIA-ALLENDE, Olga M. CONDE, Jesus MIRAPEIX, Ana M. CUBILLAS, Jose M. LOPEZ-HIGUERA

Grupo de Ingeniería Fotónica, Universidad de Cantabria, Avda. de Los Castros S/N, 39005 Santander (España)

Persona de contacto: P. Beatriz GARCIA-ALLENDE (garciapb@unican.es).

RESUMEN:

En este artículo se presenta un nuevo método de procesamiento de imágenes hiperespectrales capturadas mediante un sensor espectroscópico de imagen. Basado en un algoritmo de clasificación lineal, permite mejorar las restricciones de tiempo real que planteaba un sistema de clasificación de materia prima diseñado previamente. El procesamiento de datos consta de dos bloques: compresión de características e interpretación de las mismas. La compresión se realiza empleando Análisis de Componentes Principales (PCA), mientras que la clasificación se realiza mediante el Mapeo del Ángulo Espectral (SAM). La técnica diseñada se ha aplicado al control de la calidad de la materia prima de la industria tabacalera. En ésta se pretende separar la materia prima (las hojas de tabaco) de un conjunto de materiales diversos como madera, plásticos, cartón, papel de caramelo, etc. El sensor espectroscópico empleado para la captura de las imágenes hiperespectrales consta de una cámara monocromática y un dispositivo basado en un elemento dispersivo pasivo Prisma – Red de difracción – Prisma.

Palabras clave: Espectroscopía de absorción, imágenes hiperespectrales, Análisis de Componentes Principales (PCA), Mapeo del Ángulo Espectral (SAM), clasificación en tiempo real de materias primas.

ABSTRACT:

A data processing method for hyperspectral images from an imaging spectroscopic sensor is presented. It is a linear algorithm which comes to solve real time constraints of a previously designed raw material classification system. The new data processing method is composed of two blocks: data compression and classification unit. Data compression is performed by means of Principal Component Analysis (PCA) and the interpretation algorithm for classification is the Spectral Angle Mapper (SAM). This strategy has been successfully tested on the raw material on-line characterization in tobacco industry. In this application, the desired raw material (tobacco leaves) should be discriminated from other unwanted spurious materials, such as wood, plastic, cardboard, leather, candy paper, etc. Hyperspectral images are recorded by a spectroscopic sensor consisting of a monochromatic camera and a dispersive passive Prism-Grating-Prism device.

Key words: Absorption spectroscopy, hyperspectral images, Principal Component Analysis (PCA), Spectral Angle Mapper (SAM), raw material on-line characterization.

1.- Introducción

Se entiende por espectroscopia la monitorización de la interacción de elementos y compuestos con la luz en base a la medida de la intensidad y la energía radiante. El objeto es estudiar la estructura molecular y dinámica de los mismos. En el caso particular de la espectroscopia de absorción el elemento que interacciona con la luz es un elemento pasivo, que absorbe parte del espectro de la luz, siendo esta absorción dependiente de la longitud de onda. La radiación que no se absorbe puede transmitirse a través de la muestra del compuesto o reflejarse en ella (reflectancia difusa). A partir de la medida e interpretación del espectro de esta reflectancia difusa se pueden clasificar los diferentes compuestos o identificar los elementos químicos de los mismos.

La espectroscopia de imagen permite capturar simultáneamente las componentes ópticas del espectro y la posición espacial de un objeto en una superficie. De esta forma el espectro de la reflectancia difusa se mide al mismo tiempo en todos los píxeles de la línea de visión. Para ello se utiliza un dispositivo Prisma – Red de difracción - Prisma (PGP, *Prism-Grating-Prism*) [1]. Esta técnica óptica hiperespectral es muy adecuada para su aplicación en la discriminación y clasificación de la materia prima de diversos procesos industriales en los que la cinta transportadora se desplaza bajo el sistema de clasificación formando la segunda dimensión espacial [1-3]. En este artículo se presenta un algoritmo de interpretación y clasificación de estos datos hiperespectrales, basado en el Mapeo del Ángulo Espectral (SAM, *Spectral Angle Mapper*) [4]. La razón de seleccionar este algoritmo es que, debido a su linealidad, es muy eficiente desde el punto de vista del tiempo de ejecución. Esto permite la operación del sistema de clasificación en tiempo real, a medida que el material se desplaza sobre la cinta transportadora en la planta de producción. Las prestaciones se comparan con las Redes Neuronales (*Artificial Neural Networks, ANN*), que fue el algoritmo de interpretación espectral empleado en un diseño previo del sistema sensor [5]. Además, para facilitar la operación en tiempo real se emplea, previamente a la clasificación, un

algoritmo de extracción de la información relevante del espectro, denominada *huella espectral* del material, mediante un método basado en el procesado de autovalores y autovectores: Análisis de Componentes Principales (*PCA, Principal Component Analysis*) [6]. El tabaco ha sido empleado con éxito para demostrar experimentalmente la validez de la técnica presentada.

El artículo se divide en las secciones siguientes: en la Sección 2 se describe en detalle el procesado al que se someten las imágenes hiperespectrales, incluyendo tanto el bloque de compresión como el de interpretación. La Sección 3 presenta el montaje óptico y sus componentes. Para finalizar, en las Secciones 4 y 5, se presentan los resultados y principales conclusiones del trabajo.

2.- Procesado de imágenes hiperespectrales.

La estrategia implementada para la interpretación de datos hiperespectrales consta, como se ha mencionado anteriormente, de dos partes. En primer lugar se emplea PCA [6] para realizar de forma simultánea una eliminación de la redundancia y la reducción de la dimensionalidad de los datos. A continuación, se emplea SAM [4] para la detección de los elementos espurios, así como para la determinación de su posición espacial. En la Figura 1 se presenta un diagrama de bloques del análisis espectral propuesto.

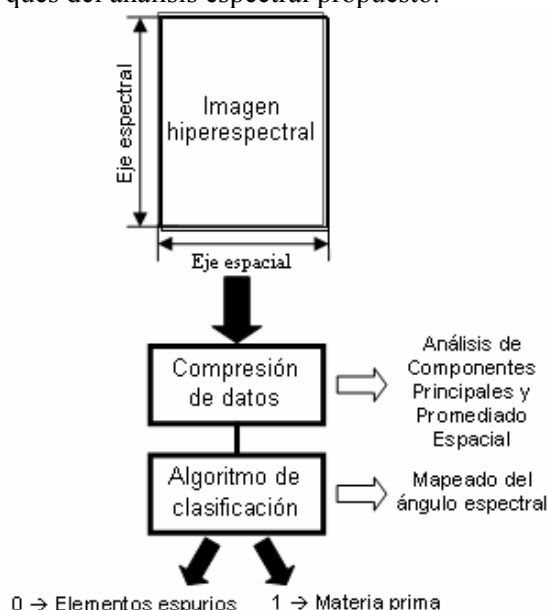


Fig. 1. Diagrama de bloques del procesado de las imágenes hiperespectrales.

2.1.- Análisis de Componentes Principales

En el bloque de compresión de datos se realiza tanto la eliminación de la redundancia del eje espectral, como un promediado del eje espacial de las imágenes. Este último se puede realizar debido que el sistema industrial de extracción mecánica de defectos tiene baja resolución. Además de la compresión, este promediado favorece la reducción del ruido inherente a la medida.

La compresión del eje espectral se realiza mediante PCA. Consiste en proyectar el espectro de la reflectancia difusa de cada píxel de la línea de visión en una base vectorial de menor dimensión. Los vectores que conforman la base del nuevo espacio vectorial son aquellas direcciones de mayor varianza de los datos iniciales. Debido a que PCA supone que varianza implica información relevante [6], se debe, en primer lugar, restar la media para cada longitud de onda. Una vez que los datos tienen media nula, se calcula la matriz de covarianza y los autovalores y autovectores de la misma. Para realizar la compresión del eje espectral se desprecian los autovalores menores de la milésima parte del máximo. De este modo, el espectro capturado de cada posición espacial (400 a 1000 nm), que inicialmente se encuentra en un espacio de dimensión 640 es comprimido a un espacio de dimensión 18 en el que se realiza la clasificación. Esta elevada tasa de compresión, aproximadamente del 99%, permite una reducción del tiempo de clasificación sin ocasionar un incremento significativo en su probabilidad de error.

2.2.- Mapeo del ángulo espectral

Se trata de un algoritmo sencillo basado en la medida de la similitud espectral entre dos espectros. Esta similitud espectral, α , se obtiene considerando cada espectro como un vector en un espacio n -dimensional del siguiente modo[4]:

$$\alpha = \cos^{-1} \left(\frac{\sum_{i=1}^n e_{1i} e_{2i}}{\left(\sum_{i=1}^n e_1^2 \right)^{1/2} \left(\sum_{i=1}^n e_2^2 \right)^{1/2}} \right)$$

donde n es el número de bandas espectrales y e_1 y e_2 son los dos espectros que se están comparando. Un ángulo α pequeño indica elevada similitud entre los espectros, mientras que ángulos elevados indican baja similitud. Debido a la aplicación previa de PCA, ahora el eje espectral tiene solamente 18 bandas.

La implementación del método de clasificación basado en SAM se compone de las dos partes siguientes:

1. En primer lugar, en la etapa de *entrenamiento*, se deben obtener espectros de referencia tanto de la materia prima, e_{mp} , deseada como de los elementos espurios, e_d .

Para ello se han medido 30 imágenes *homogéneas* de hojas de tabaco (2880 espectros) y 24 de los defectos que típicamente suelen aparecer en esta aplicación (2304 espectros). En este caso “homogéneo” quiere decir que el mismo tipo de material aparece a lo largo de toda la línea de visión.

2. En la etapa de *test*, se clasifica una imagen desconocida. Para ello se calculan las similitudes espectrales α_{t-mp} (entre los espectros de cada una de las posiciones espaciales de la línea de imagen de test, e_t , y el espectro de referencia de la materia prima deseada, el de la hoja de tabaco) y α_{t-d} (entre los espectros de test y el espectro de referencia de los materiales espurios o defectos). Si $\alpha_{t-mp} < \alpha_{t-d}$ el espectro de test,

e_t , está más próximo al espectro de referencia de la materia prima y por tanto esa posición espacial se corresponde con tabaco. Si por el contrario $\alpha_{t-d} < \alpha_{t-mp}$ indica que e_t se corresponde con una posición espacial en la que hay un material espurio o “defecto”. En la Figura 2 se presenta un esquema de la estrategia de clasificación propuesta.

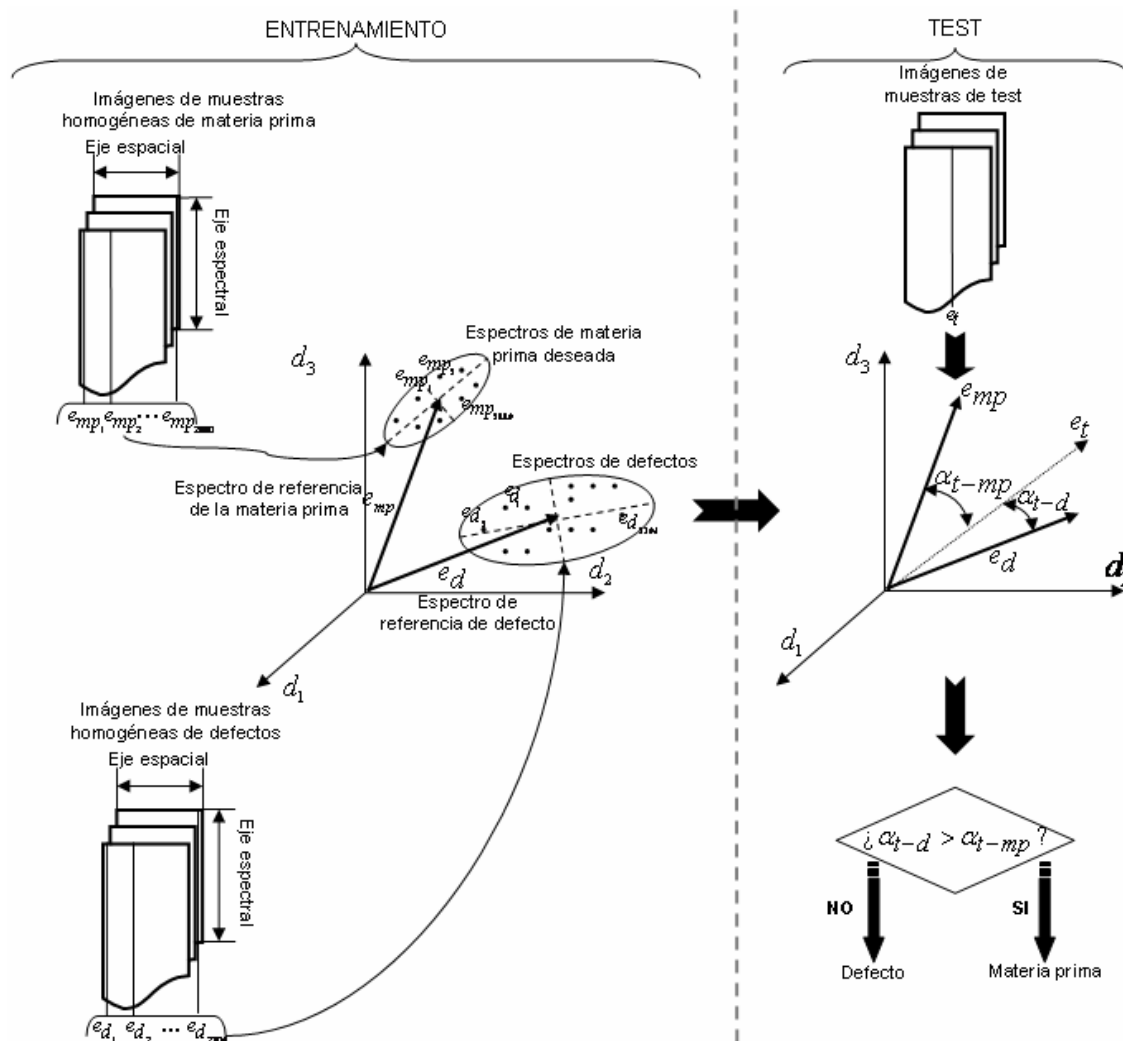


Fig.2. Descripción esquemática del algoritmo de interpretación espectral (SAM).

3.- Montaje experimental.

La adquisición de imágenes espectrales está basada en un espectrógrafo de imagen PGP. Es un elemento totalmente pasivo conectado entre una lente de enfoque y una cámara equipada con el interfaz IEEE1394. En la Figura 3 se presenta un diagrama de bloques de este montaje.

Como sistema de iluminación se han empleado dos focos halógenos de 500W de potencia cada uno.

El espectrógrafo de imagen empleado ha sido el equipo comercial denominado ImS-pector en su versión V10E, cuyo rango espectral va desde 400 a 1000nm.

La selección de la cámara depende de la aplicación. Además, el ImSpector está diseñado para detectores estándar de 2/3" como es el de la cámara monocromática Pixelink PL-A741 MV que aquí se ha empleado.

Por último, la lente de enfoque elegida ha sido la Zoom 7000 de Navitar.

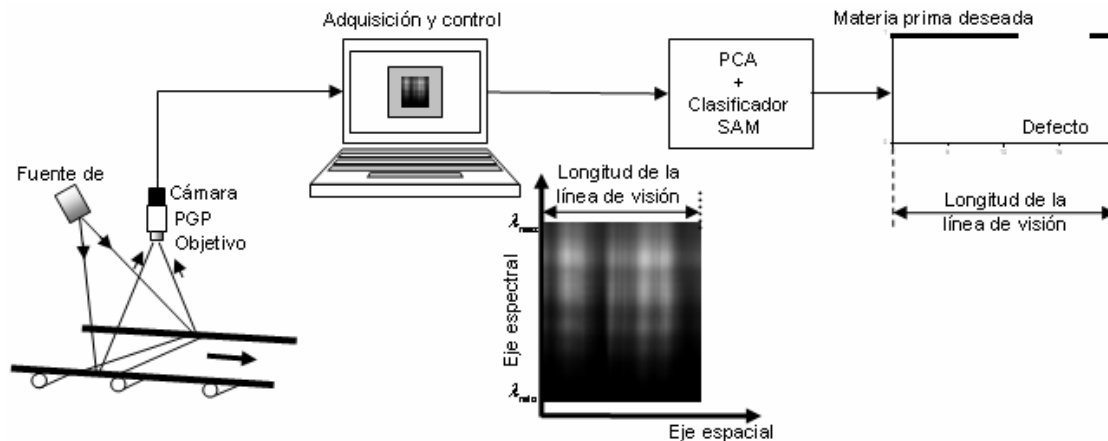


Fig.3. Diagrama de bloques del montaje experimental.

4.- Resultados y discusión.

Como ya se ha mencionado, la técnica desarrollada se ha validado con éxito con productos de la industria tabacalera. Las hojas de tabaco utilizadas fueron proporcionadas por la compañía Altadis y junto a ellas se consideraron diversos materiales para ser discriminados de éstas: madera, cartón, plásticos de diversos colores, cuero, hojas de otros vegetales, hilo, papel de aluminio y papel de caramelo. La Tabla 1 resume e ilustra los resultados obtenidos. Recoge en filas la siguiente información:

1. Fotografía de la muestra analizada, en la que las líneas blancas delimitan la línea iluminada. Después del promediado espacial se tienen 96 píxeles diferentes en esa línea.
2. Zoom en la región de la imagen capturada por el sistema.
3. Las imágenes hiperespectrales. El eje horizontal es el eje espacial con el origen en la izquierda, al igual que en las fotografías reales. El eje vertical es el eje espectral con la longitud de onda mínima en la parte inferior.
4. Resultados de clasificación proporcionados por el algoritmo SAM. Si en una posición espacial particular se tiene un '1' esto indica que se corresponde con tabaco y si aparece un '0' indica la existencia de un defecto en ese punto.

La muestra analizada en primer lugar (columna A) presenta un taco de madera sobre las hojas de tabaco. La diferencia se puede ver claramente en la imagen hiperespectral y por último como la salida del sis-

tema de clasificación indica la presencia de un defecto hasta $x \approx 12$ mm.

A continuación, en la columna B, se presenta una muestra con un trozo de celofán amarillo situado encima de las hojas de tabaco. De nuevo se puede apreciar, también claramente, la presencia de este material espurio en la imagen hiperespectral y cómo el sistema identifica adecuadamente la presencia de un defecto en la región espacial que va de $x \approx 11$ mm hasta $x \approx 14.5$ mm.

Por último, y con el objetivo de demostrar que la técnica de clasificación no está basada en el análisis colorimétrico, se presenta en la columna C una muestra que consta de una hoja de otro material vegetal diferente de tabaco mezclada con hojas de tabaco. Se ve claramente como el sistema detecta la presencia de un material espurio en la región espacial comprendida entre $x \approx 2.5$ mm y $x \approx 11.5$ mm. En este caso se detectado también una "falsa alarma", esto es, un espectro de hoja de tabaco clasificado como defecto, en la interfaz entre las primeras hojas de tabaco y el inicio de la hoja que no es de tabaco. Esto ocasiona lógicamente un incremento de la probabilidad de error pero es admisible comparado con la no detección de un defecto. En definitiva sobre un total de 32 imágenes de test (3072 espectros) se ha obtenido un porcentaje de acierto en la clasificación del 90.54%.

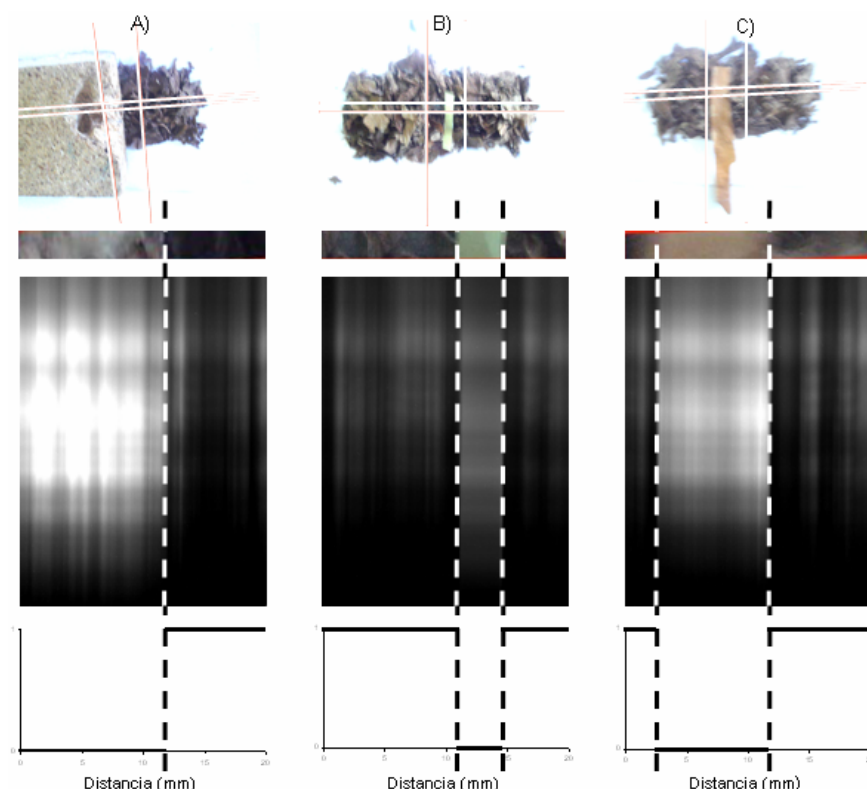


Tabla 1. Resultados de la clasificación.

5.- Conclusiones.

En este trabajo se ha diseñado un método de procesamiento de datos hiperespectrales que consiste en la aplicación correlativa del Análisis de Componentes Principales y el Mapeado Espectral. Esta estrategia se ha aplicado con éxito al control de calidad, en tiempo real, de la materia prima de la industria tabacalera. El empleo de PCA permite mejorar las prestaciones temporales del SAM. En la Tabla 1 se puede ver como el algoritmo realiza una discriminación eficiente de los diferentes defectos. Esta eficiencia reside en la simplicidad y linealidad del algoritmo SAM.

Agradecimientos

Este trabajo ha sido en parte realizado con los medios aportados por los proyectos CICYT TEC2004-05936-C02-02/MIC y TEC2005-08218-C02-02.

Referencias

- [1] Takio VAARALA, Mauri AIKIO y Heimo KERAENEN, "An advanced prism-grating-prism imaging spectrograph in on-line industrial applications", Proceedings del SPIE, Volume 3101, 322-30, 1997.
- [2] Esko HERRALA, Timo HYVARINEN, Osmo VOUTILAINEN y Jorma

LAMMASNIEMI, "An optoelectronic sensor system for industrial multipoint and imaging spectrometry", Sensors and Actuators A (Physical), Volume A61, n 1-3, 335-338, 1997.

- [3] Diwan P. ARIANA, Renfu LU, Daniel E. GUYER, "Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers", Computers and Electronics in Agriculture, Volume 53, n 1, 60-70, 2006.
- [4] G GIROUARD, A BANNARI, A EL HARTI, A DESROCHERS, "Validated Spectral Angle Mapper Algorithm for Geological Mapping: Comparative Study between Quickbird and Landsat-TM", Geo-Imagery Bridging Continents, Istanbul, July 12-23, 2004, 599-604.
- [5] P. Beatriz GARCÍA-ALLENDE, Olga M CONDE, Ana M. CUBILLAS, César JÁUREGUI and José M. LÓPEZ-HIGUERA, "New Raw Material Discrimination System Based on a Spatial Optical Spectroscopy Technique", Sensors and Actuators A(Physical) In Print.
- [6] J WORKMAN Jr, A W. SPRINGSTEEN, "Applied Spectroscopy: A compact reference for Practitioners", first ed., Academic Press Limited, London, 1998.